

# HYPERSPECTRAL SHARPENING USING SCENE-ADAPTED GAUSSIAN MIXTURE PRIORS

Afonso M. Teodoro   José M. Bioucas-Dias   Mário A. T. Figueiredo

Instituto de Telecomunicações  
Instituto Superior Técnico, Universidade de Lisboa, Portugal

## ABSTRACT

This paper tackles a hyperspectral data fusion problem, using the so-called plug-and-play approach, which combines an ADMM algorithm with a denoiser based on a Gaussian mixture prior. We build upon the concept of scene-adapted prior where, as the name suggests, we learn a model that is targeted to the specific scene being imaged, and show state-of-the-art results on several hyperspectral sharpening experiments. Additionally, we prove that the algorithm is guaranteed to converge.

**Index Terms**— Data fusion, hyperspectral sharpening, plug-and-play, Gaussian mixture, scene-adapted prior.

## 1. INTRODUCTION

Recently, we have proposed using class-adapted Gaussian mixture priors to address several inverse imaging problems, such as deblurring and compressive imaging [14, 15], by exploiting the so-called *plug-and-play* (PnP) approach [17]. We proposed training a *Gaussian mixture model* (GMM) from external datasets containing images of particular classes (such as text, faces, fingerprints), and then *plug* a GMM-based denoiser [13] into the iterations of an *alternating direction method of multipliers* (ADMM) [6], obtaining state-of-the-art results on images of the specific classes considered.

We have recently shown that the same machinery can be used to address a different task: hyperspectral image sharpening (also known as fusion) [16]. In this problem, the goal is to fuse two types of data describing the same scene, thus we can use one of them to train a model that is targeted to that scene. Under the hypothesis that both data have the same spatial structure (since they are imaging the same scene), we can leverage this model to reconstruct the other type of data. Hence the designation *scene-adapted* prior.

This paper differs from [16] in two major aspects. First, following the same rationale as in [18], we propose a modification of the GMM denoiser used in [16], which can be interpreted as keeping the dictionary support fixed throughout the iterations of ADMM. Second, although we observed convergence in practice in [16], this modification enables us to theoretically guarantee convergence.

The paper is organized as follows. Section 2 formulates the addressed problem while Section 3 shows how to tackle it using a PnP-ADMM scheme. Sections 4 and 5 describe the adopted denoiser and show that the proposed PnP-ADMM is guaranteed to converge to a minimum of an implicitly defined objective function. In Section 6, we present experimental results showing that the proposed method performs competitively with, in some cases better than, another state-of-the-art method. Section 7 concludes the paper.

This work was partially supported by the *Fundação para a Ciência e Tecnologia* (FCT), grants UID/EEA/5008/2013, BD/102715/2014, and PTDC/EEI-PRO/0426/2014.

## 2. PROBLEM FORMULATION

As mentioned above, hyperspectral image sharpening is a data fusion problem where the goal is to combine two types of observed data: a *hyperspectral* (HS) data cube, with very high spectral resolution, but low spatial resolution; a *multispectral* (MS) data cube, with high spatial resolution, but low spectral resolution. The observation model underlying this data fusion problem assumes that the HS bands are blurred and down-sampled versions of a target cube with high spectral and spatial resolution, whereas the MS bands are spectrally degraded versions of the same target cube, due to the spectral responses of the sensor(s). The forward model is thus

$$\mathbf{Y}_h = \mathbf{Z}\mathbf{B} + \mathbf{N}_h, \quad (1)$$

$$\mathbf{Y}_m = \mathbf{R}\mathbf{Z} + \mathbf{N}_m, \quad (2)$$

where  $\mathbf{Z} \in \mathbb{R}^{L_h \times n_m}$  is the target cube to be estimated (with rows indexing bands, and columns indexing spatial locations),  $\mathbf{Y}_h \in \mathbb{R}^{L_h \times n_h}$  is the observed HS data,  $\mathbf{B} \in \mathbb{R}^{n_m \times n_h}$  models a spatial convolution,  $\mathbf{M} \in \mathbb{R}^{n_m \times n_h}$  is a sub-sampling operator,  $\mathbf{Y}_m \in \mathbb{R}^{L_m \times n_m}$  is the observed MS data,  $\mathbf{R} \in \mathbb{R}^{L_m \times L_h}$  models the spectral responses of the MS sensor, and  $\mathbf{N}_h$  and  $\mathbf{N}_m$  are Gaussian noises of known variance. For simplicity, we assume that the sensor has an invariant point spread function across the  $L_h$  spectral bands and the matrices  $\mathbf{B}$  and  $\mathbf{R}$  are known. Estimation of those matrices is addressed in [10].

Another common feature of HS data is that it lives, with very good approximation, in a subspace of lower dimension, say  $k$ , than the number of bands  $L_h$  [4], and this subspace can be accurately inferred from the HS data cube [5]. Therefore, we may write  $\mathbf{Z} = \mathbf{E}\mathbf{X}$ , where  $\mathbf{E} \in \mathbb{R}^{L_h \times k}$ . Instead of estimating  $\mathbf{Z}$  directly, we infer  $\mathbf{X}$ , denoted latent or coefficient image, and then recover the target image. With this simplification, the forward model becomes

$$\mathbf{Y}_h = \mathbf{E}\mathbf{X}\mathbf{B} + \mathbf{N}_h, \quad (3)$$

$$\mathbf{Y}_m = \mathbf{R}\mathbf{E}\mathbf{X} + \mathbf{N}_m. \quad (4)$$

To the best of our knowledge, the current best methods for hyperspectral fusion seek a *maximum a posteriori* (or regularized) estimate of  $\mathbf{X}$  as

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{E}\mathbf{X}\mathbf{B} - \mathbf{Y}_h\|_F^2 + \lambda \|\mathbf{R}\mathbf{E}\mathbf{X} - \mathbf{Y}_m\|_F^2 + 2\tau \phi(\mathbf{X}), \quad (5)$$

where  $\phi$  is the negative log-prior (or regularizer), while  $\lambda$  and  $\tau$  control the relative weight of each term. The regularizer can be chosen to induce spatial coherence (e.g., via TV regularization [10]) or some form of sparse representation on a dictionary [18].

In (5), we implicitly assumed that the components of the noise matrices  $\mathbf{N}_h$  and  $\mathbf{N}_m$  are independent and Gaussian, with zero-mean and constant variance. Other forms of noise correlations, decoupled with respect to the bands and to the pixels, are easily accommodated in our formulation (see, for example, [18]).

### 3. PNP ADMM

The tool of choice to solve (5) has been ADMM [1, 6]. Due to space limitations, we do not describe the derivations leading to this instance of ADMM (for details, see [10, 18] and references therein). In summary, ADMM solves a sequence of sub-problems w.r.t. each of the primal variables, while keeping the others fixed, followed by updating the dual variables. This process is repeated until some criterion is satisfied. The ADMM algorithm that solves (5) has the form

$$\begin{aligned} \mathbf{X}^{k+1} &= \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{B} - \mathbf{V}_1 - \mathbf{D}_1\|_F^2 + \|\mathbf{X} - \mathbf{V}_2 - \mathbf{D}_2\|_F^2 + \|\mathbf{X} - \mathbf{V}_3 - \mathbf{D}_3\|_F^2, \\ \mathbf{V}_1^{k+1} &= \underset{\mathbf{V}_1}{\operatorname{argmin}} \|\mathbf{E}\mathbf{V}_1\mathbf{M} - \mathbf{Y}_h\|_F^2 + \rho\|\mathbf{X}^{k+1}\mathbf{B} - \mathbf{V}_1 - \mathbf{D}_1^k\|_F^2, \\ \mathbf{V}_2^{k+1} &= \underset{\mathbf{V}_2}{\operatorname{argmin}} \lambda\|\mathbf{R}\mathbf{E}\mathbf{V}_2 - \mathbf{Y}_m\|_F^2 + \rho\|\mathbf{X}^{k+1} - \mathbf{V}_2 - \mathbf{D}_2^k\|_F^2, \\ \mathbf{V}_3^{k+1} &= \underset{\mathbf{V}_3}{\operatorname{argmin}} 2\tau\phi(\mathbf{V}_3) + \rho\|\mathbf{X}^{k+1} - \mathbf{V}_3 - \mathbf{D}_3^k\|_F^2, \end{aligned}$$

where  $\mathbf{X}, \mathbf{V}_1, \mathbf{V}_2$ , and  $\mathbf{V}_3$  are the primal variables,  $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  are the scaled dual variables, and  $\rho$  is a penalty parameter. The first three problems are quadratic, with closed-form solutions,

$$\begin{aligned} \mathbf{X}^{k+1} &= \left[ \left( \mathbf{V}_1^k + \mathbf{D}_1^k \right) \mathbf{B}^T + \left( \mathbf{V}_2^k + \mathbf{D}_2^k \right) + \left( \mathbf{V}_3^k + \mathbf{D}_3^k \right) \right] \left[ \mathbf{B}\mathbf{B}^T + 2\mathbf{I} \right]^{-1}, \\ \mathbf{V}_1^{k+1} &= \left[ \mathbf{E}\mathbf{E}^T + \rho\mathbf{I} \right]^{-1} \left[ \mathbf{E}^T\mathbf{Y}_h + \rho \left( \mathbf{X}^{k+1}\mathbf{B} - \mathbf{D}_1^k \right) \right] \odot \mathbf{M} \\ &\quad + \left( \mathbf{X}^{k+1}\mathbf{B} - \mathbf{D}_1^k \right) \odot (1 - \mathbf{M}), \\ \mathbf{V}_2^{k+1} &= \left[ \lambda\mathbf{E}^T\mathbf{R}^T\mathbf{R}\mathbf{E} + \rho\mathbf{I} \right]^{-1} \left[ \lambda\mathbf{E}^T\mathbf{R}^T\mathbf{Y}_m + \rho \left( \mathbf{X}^{k+1} - \mathbf{D}_2^k \right) \right], \end{aligned}$$

where the symbol  $\odot$  denotes entry-wise (Hadamard) product. The optimization with respect to  $\mathbf{V}_3^{k+1}$  corresponds to the *proximity operator* of the regularizer  $\phi$  [2] (which can be seen as a denoising operator) acting on  $(\mathbf{X}^{k+1} - \mathbf{D}_3^k)$ . The usual choice for  $\phi$  is a convex regularizer, for two main reasons: computational tractability and convergence guarantees for ADMM [6]. The fact that state-of-the-art denoisers (such as the famous BM3D [8]) cannot, in general, be formulated as the proximity operators of convex regularizers has motivated the PnP approach [17], where the proximity operator is replaced with a black-box denoising algorithm.

In this paper, we adopt a PnP approach using a patch-based *minimum mean squared error* (MMSE) denoiser with a GMM prior [13, 19, 20]. A GMM prior can either be learned directly from noisy image patches [13] or from an external set of clean patches [20]. In particular, this latter possibility motivated our class-adapted GMM priors [14, 15], where the external set contains images from of a specific class (or classes). Here, we take this GMM-based prior adaptation one step further, by considering a scene-adapted prior. Specifically, we use a GMM prior learned from the MS images (the bands in  $\mathbf{Y}_m$ ), under the assumption that they have the same spatial statistical properties as the underlying unknown image, since they are acquired from the same scene.

### 4. GMM-BASED DENOISER

A key fact motivating the use of GMM priors is that the corresponding MMSE estimate, from observations with Gaussian noise, can be computed in closed-form [13]. Letting  $\mathbf{x}_i$  and  $\mathbf{y}_i$  be an arbitrary patch of the unknown clean image and a noisy version thereof (zero-mean Gaussian noise, with variance  $\sigma^2$ ), the MMSE estimate of  $\hat{\mathbf{x}}_i$  is

$$\hat{\mathbf{x}}_i = \mathbb{E}[\mathbf{x}_i|\mathbf{y}_i] = \sum_{m=1}^K \beta_m(\mathbf{y}_i) \mathbf{v}_m(\mathbf{y}_i), \quad (6)$$

where

$$\mathbf{v}_m(\mathbf{y}_i) = \mathbf{C}_m \left( \mathbf{C}_m + \sigma^2 \mathbf{I} \right)^{-1} \left( \sigma^2 \mathbf{C}_m^{-1} \mu_m + \mathbf{y}_i \right), \quad (7)$$

and

$$\beta_m(\mathbf{y}_i) = \frac{\alpha_m \mathcal{N}(\mathbf{y}_i; \mu_m, \mathbf{C}_m + \sigma^2 \mathbf{I})}{\sum_{j=1}^K \alpha_j \mathcal{N}(\mathbf{y}_i; \mu_j, \mathbf{C}_m + \sigma^2 \mathbf{I})}. \quad (8)$$

In Eqs. (6)-(8),  $K$  is the number of mixture components,  $\mathcal{N}(\cdot; \mu, \mathbf{C})$  denotes a Gaussian probability density function of mean  $\mu$  and covariance  $\mathbf{C}$ , and  $\alpha_m$  is the weight of the  $m$ -th mixture component.

For simplicity, we use only zero-mean Gaussian components (thus a zero-mean GMM), i.e.,  $\mu_m = 0$ , for  $m = 1, \dots, K$ . This is reasonable since the variance of the noise affecting the mean of each patch is only  $\sigma^2/n_p$ , where  $n_p$  is the number of pixels in each patch. Hence, we can remove the mean of each noisy patch, obtain the MMSE estimate, and add back the mean.

The final task of the denoiser is to combine the overlapping patch estimates  $\hat{\mathbf{x}}_i$  to obtain a final image  $\hat{\mathbf{x}}$ , which can be formulated as an optimization problem,

$$\hat{\mathbf{x}} \in \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{P}_i \mathbf{x}\|_2^2, \quad (9)$$

where  $\mathbf{P}_i \in \{0, 1\}^{n_p \times n_m}$  is a binary matrix that extracts the  $i$ -th patch from the image (thus  $\mathbf{P}_i^T$  puts the patch back into its place) and  $N$  is the number of patches. The solution to (9) is

$$\hat{\mathbf{x}} = \left( \sum_{i=1}^N \mathbf{P}_i^T \mathbf{P}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{P}_i^T \hat{\mathbf{x}}_i \right). \quad (10)$$

Assuming that the patches are extracted with unit step and periodic boundary conditions, every pixel belongs to  $n_p$  patches and  $\sum_{i=1}^N \mathbf{P}_i^T \mathbf{P}_i = n_p \mathbf{I}$ . In this case, (10) corresponds to combining the overlapping estimates of each pixel using a straight average. Extension to other patch extraction schemes is straightforward; in any case, as long as all pixels belong to at least to one patch, the matrix inversion in (10) is well defined as yields a diagonal matrix.

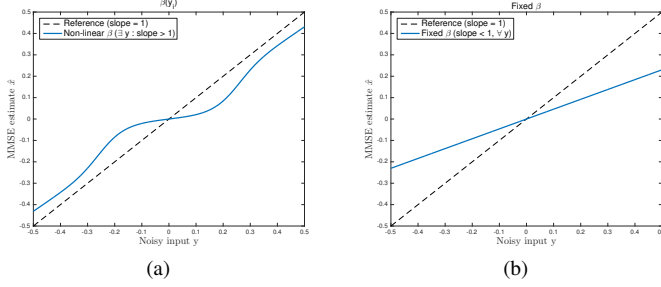
### 5. CONVERGENCE OF THE PNP-ADMM ALGORITHM

Having chosen a denoiser to plug into ADMM, we turn our attentions to the convergence of the resulting algorithm. In [11], the authors provide two sufficient conditions on the denoiser that ensure convergence to a global minimum of some implicitly defined objective function: it should be non-expansive<sup>1</sup> and the sub-gradient of a convex function. In practice, these conditions are restrictive and hard to verify by a black-box denoiser.

The GMM denoiser described in Section 4 is not non-expansive, as shown in a simple one-dimensional counter-example in Fig. 1a, which plots the MMSE estimate  $\hat{x}$  in (6), as a function of the noisy observed  $y$ , under a GMM prior with 2 zero mean components. Clearly, the presence of regions of the function with derivative larger than 1 shows that it is not non-expansive. Note that this does not invalidate the convergence of the PnP-ADMM with the GMM-denoiser; in fact, we observe convergence in practice and, under some mild conditions, such as those proposed in [7], we can guarantee that the iterates converge to a fixed point.

<sup>1</sup>Recall that a function  $\mathcal{D}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is *non-expansive* if, for any  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,

$$\|\mathcal{D}(\mathbf{x}) - \mathcal{D}(\mathbf{y})\|^2 \leq \|\mathbf{x} - \mathbf{y}\|^2. \quad (11)$$



**Fig. 1:** Denoiser expansiveness example: (a) Non-linear weights,  $\beta(y_i)$ ; (b) Fixed weights,  $\beta$ .

The failure to be non-expansive of the MMSE denoiser can be traced to the non-linearity of the weights  $\beta_m(y)$ . In fact, each function  $\mathbf{v}_m$  defined in (7) is itself non-expansive, because the eigenvalues of  $(\sigma^2 \mathbf{C}_m^{-1} + \mathbf{I})^{-1}$  are no larger than 1; thus, if the weights are fixed ( $\beta_m \geq 0$ ,  $\sum_m \beta_m = 1$ ), the corresponding convex combination of non-expansive operators is non-expansive [2], as illustrated with a one-dimensional example in Fig. 1b.

We claim that, if the hyperspectral fusion problem in Eqs. (3)–(4) is addressed with a GMM prior learned from the MS images, it makes sense to keep the weights  $\beta_m$  of each patch equal to its value for the corresponding patch of  $\mathbf{Y}_m$ . In other words, we use the weights as if we were denoising the high resolution MS image, with a GMM trained from its noisy patches. In fact, another instance of this idea was previously used in [18]. There, the authors propose an approach to the same problem, based on sparse representations of the patches on a dictionary learned from the MS images, but instead of recomputing the sparse support of the dictionary and the corresponding weights, the support is kept fixed and equal to the support of the representation of the corresponding patch in the MS images. We stress that this approach is only reasonable in the case of image denoising or image reconstruction using scene-adapted priors, where the training and target images are of the same scene, and thus have a similar spatial structure.

Recalling that  $\mu_m = 0$ , for  $m = 1, \dots, K$ , the patch estimates are

$$\hat{\mathbf{x}}_i = \sum_{m=1}^K \beta_m^i \mathbf{C}_m (\mathbf{C}_m + \sigma^2 \mathbf{I})^{-1} \mathbf{y}_i = \mathbf{F}_i \mathbf{y}_i = \mathbf{F}_i \mathbf{P}_i \mathbf{y}, \quad (12)$$

with  $\beta_m^i$  and  $\mathbf{C}_m$ , for  $m = 1, \dots, K$ , computed at the training stage, and kept constant thereafter. Note that although the weights  $\beta_m^i$  no longer depend on the values in patch  $\mathbf{y}_i$ , we introduce a superscript to denote that each patch has a different set of fixed weights  $\beta_m^i \geq 0$ , such that  $\sum_m \beta_m^i = 1$ . The second equality uses matrix  $\mathbf{P}_i$ , which extracts patch  $\mathbf{y}_i$  from  $\mathbf{y}$ . Plugging Eq. (12) into (10) yields

$$\hat{\mathbf{x}} = \frac{1}{n_p} \sum_{i=1}^N \mathbf{P}_i^T \mathbf{F}_i \mathbf{P}_i \mathbf{y} = \mathbf{W}_{GMM} \mathbf{y}, \quad (13)$$

showing that the denoised image is a linear function of the noisy one. Clearly,  $\mathbf{W}_{GMM}$  is symmetric, as is obvious from its definition and the fact that each  $\mathbf{F}_i$  is symmetric because it is a convex combination of symmetric matrices.

In [12], a very similar formula was presented for approximate MAP estimation under a GMM prior, where the authors include a parameter to avoid numerical instability when the noise level is very small. Moreover, Theorem 1 in [12] remains valid here. Recall that the weights  $\beta_m^i$ , although fixed, are non-negative and sum to one, and

that the covariance matrices of the GMM components are positive-semidefinite (p.s.d.). In fact, since we train the model using a low-noise panchromatic image (as will be seen in the next section), most, if not all, singular values of each  $\mathbf{C}_m$  are strictly positive; besides, adding a small constant to the diagonal to ensure this holds does not change the obtained results whatsoever. Hence, every matrix  $\mathbf{F}_i$  (defined in Eq. (12)) is a convex combination of positive-definite (p.d.) matrices, thus also a p.d. matrix.

Using the eigendecomposition  $\mathbf{C}_m = \mathbf{U}_m^T \mathbf{S}_m \mathbf{U}_m$  of each covariance matrix, where  $\mathbf{S}_m = \text{diag}(s_1^m, \dots, s_{n_p}^m)$  contains the eigenvalues of  $\mathbf{C}_m$ , sorted in non-increasing order, we may write

$$\begin{aligned} \mathbf{C}_m (\mathbf{C}_m + \sigma^2 \mathbf{I})^{-1} &= \mathbf{U}_m^T \mathbf{S}_m \mathbf{U}_m (\mathbf{U}_m^T \mathbf{S}_m \mathbf{U}_m + \sigma^2 \mathbf{U}_m^T \mathbf{U}_m)^{-1} \\ &= \mathbf{U}_m^T \mathbf{S}_m (\mathbf{S}_m + \sigma^2 \mathbf{I})^{-1} \mathbf{U}_m, \end{aligned} \quad (14)$$

where  $\mathbf{S}_m (\mathbf{S}_m + \sigma^2 \mathbf{I})^{-1}$  is a diagonal matrix. Consequently, all the eigenvalues of  $\mathbf{C}_m (\mathbf{C}_m + \sigma^2 \mathbf{I})^{-1}$  are in

$$\left[ \frac{s_{n_p}^m}{s_{n_p}^m + \sigma^2}, \frac{s_1^m}{s_1^m + \sigma^2} \right] \in (0, 1),$$

since  $s_{n_p}^m > 0$  and  $\sigma^2 > 0$ . From (12), each  $\mathbf{F}_i$  is a convex combination

$$\mathbf{F}_i = \sum_{m=1}^K \beta_m^i \mathbf{C}_m (\mathbf{C}_m + \sigma^2 \mathbf{I})^{-1}; \quad (15)$$

Weyl's inequality [3] implies that the eigenvalues of a convex combination of symmetric matrices is bounded below (above) by the same convex combination of the smallest (largest) eigenvalues of those matrices. Consequently, the eigenvalues of  $\mathbf{F}_i$  are all in  $(0, 1)$ , that is,  $\mathbf{F}_i$  is a p.d. matrix and  $\|\mathbf{F}_i\|_2 < 1$ .

We now show that the same is true for  $\mathbf{W}_{GMM}$ . Consider the following decomposition:

$$\mathbf{W}_{GMM} = \frac{1}{n_p} \sum_{i=1}^N \mathbf{P}_i^T \mathbf{F}_i \mathbf{P}_i = \frac{1}{n_p} \sum_j \sum_{k \in \Omega_j} \mathbf{P}_k^T \mathbf{F}_k \mathbf{P}_k = \frac{1}{n_p} \sum_j \mathbf{A}_j, \quad (16)$$

where  $\Omega_j$  denotes a set of non-overlapping patches, and the sum over  $j$  accounts for all possible shifts of the image, as in [12]. When considering non-overlapping patches, there is a permutation of the image pixels that allows writing  $\mathbf{A}_j$  as a block-diagonal matrix, where the blocks are the  $\mathbf{F}_k$  matrices, with  $k \in \Omega_j$ . Since the set of eigenvalues of a block-diagonal matrix is the union of the sets of eigenvalues of its blocks, the eigenvalues of each  $\mathbf{A}_j$  are within the same bounds as the eigenvalues of the  $\mathbf{F}_k$  matrices, i.e., they are all in  $(0, 1)$ . Finally, invoking Weyl's inequality once again, we can conclude the eigenvalues of  $\mathbf{W}_{GMM}$  are bounded above (below) by the average of the largest (smallest) eigenvalues of each  $\mathbf{A}_j$ , thus also in  $(0, 1)$ .

Summarizing, we have shown that the GMM denoiser, with the proposed modification (fixed weights), is a non-expansive operator. Non-expansiveness is one of the sufficient conditions stated in [11] for the convergence of PnP-ADMM. Following the same reasoning as in [11], we can also prove that the denoiser is the sub-gradient of some convex function, and thus it is a proximity operator [2], ensuring that ADMM converges to a global minimum, if one exists. In fact, we can guarantee that a minimum does exist because the denoiser described above is contractive:  $\|\mathbf{W}_{GMM}\|_2 < 1$ . Recall that the proximity operator of  $\phi : \mathbb{R} \rightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$  is defined as

$$\text{prox}_\phi(\mathbf{y}) = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \phi(\mathbf{x}). \quad (17)$$

**Table 1:** HS and MS fusion on (cropped) ROSIS Pavia University and Moffett Field datasets.

Dataset	Metric	Exp. 1 (PAN)			Exp. 2 (PAN)			Exp. 3 (R,G,B,N-IR)			Exp. 4 (R,G,B,N-IR)		
		ERGAS	SAM	SRE	ERGAS	SAM	SRE	ERGAS	SAM	SRE	ERGAS	SAM	SRE
Rosis	Dictionary [18]	1.99	3.28	22.64	2.05	3.16	22.32	<b>0.47</b>	<b>0.85</b>	<b>34.60</b>	0.85	1.47	29.66
	GMM [16]	1.75	2.89	23.67	1.92	2.92	22.85	0.48	0.87	34.32	0.91	1.65	29.05
	<b>proposed GMM</b>	<b>1.65</b>	<b>2.75</b>	<b>24.17</b>	<b>1.81</b>	<b>2.76</b>	<b>23.31</b>	0.49	0.87	34.59	<b>0.80</b>	<b>1.42</b>	<b>30.14</b>
Moffett	Dictionary [18]	2.67	4.18	20.28	2.74	4.20	20.05	1.85	2.72	23.58	2.12	3.21	22.25
	GMM [16]	2.66	4.24	20.26	2.78	4.27	19.87	1.81	2.68	23.81	1.98	2.93	22.91
	<b>proposed GMM</b>	<b>2.54</b>	<b>4.06</b>	<b>20.66</b>	<b>2.65</b>	<b>4.10</b>	<b>20.28</b>	<b>1.73</b>	<b>2.58</b>	<b>24.18</b>	<b>1.97</b>	<b>2.90</b>	<b>22.94</b>

From (13), we see that the denoiser is linear w.r.t. to  $\mathbf{y}$ , so a possible regularizer is  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x}$ . In fact, if  $\phi(\mathbf{x}) = \mathbf{x}^T \mathbf{Q} \mathbf{x}$ , the proximity operator becomes

$$\text{prox}_{\phi}(\mathbf{y}) = (\mathbf{I} + \mathbf{Q})^{-1} \mathbf{y}; \quad (18)$$

comparing to (13), we obtain  $\mathbf{Q} = \mathbf{W}_{GMM}^{-1} - \mathbf{I}$ , where the inverse is well defined since  $\mathbf{W}_{GMM} \succ 0$ . Moreover, since  $\|\mathbf{W}_{GMM}\|_2 < 1$ , matrix  $\mathbf{Q}$  is also p.d., thus  $\phi$  is a coercive function, ensuring that a solution to (5) exists [2].

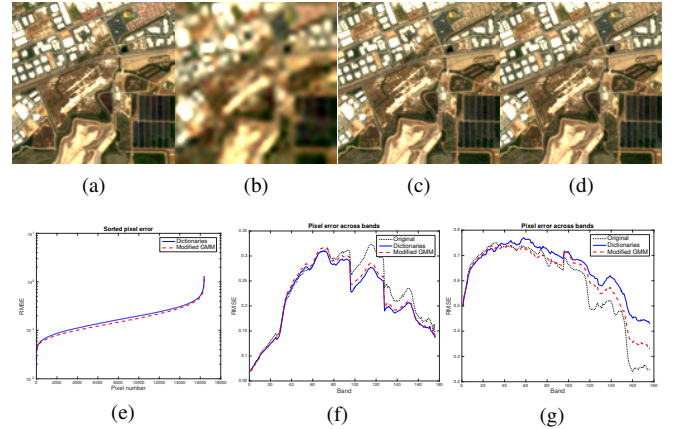
Finally, for the convergence of ADMM to be guaranteed, we must consider that the denoiser is always the same, thus parameter  $\sigma^2$  must be kept constant. In practice, this is not restrictive as long as we consider that, after a number of iterations,  $\sigma^2$  converges to some value. The first few iterations can then be seen as a transient stage, which in fact only changes the initial estimate of the algorithm.

## 6. RESULTS

In this section, we compare the hyperspectral sharpening results using the GMM-based denoiser from [13], the GMM denoiser herein proposed, and the dictionary-based method from [18] (which, to the best of our knowledge, holds the state-of-the-art). We measured the performance of all of the above using three different metrics, namely ERGAS (*erreur relative globale adimensionnelle de synthèse*), SAM (*spectral angle mapper*), and SRE (*signal to reconstruction-error*) [9], in four different settings. In the first experiment, we considered sharpening the hyperspectral images using the panchromatic (PAN) image, at a signal-to-noise ratio (SNR) of 50dB, on both the HS and PAN images. The second experiment refers to PAN-sharpening as well, this time with 35dB SNR on the first 43 hyperspectral bands and 30dB on the 50 remaining HS bands and PAN image, as in [18]. Experiments 3 and 4 consider the same noise levels as in experiments 1 and 2, respectively, but the HS bands are sharpened based on 4 MS bands, namely the R, G, B, and near-infrared channels. In every experiment, a GMM with 20 components was learned from the PAN image, where the PAN and MS images were generated from the original HS bands, using IKONOS and LANDSAT spectral responses.

Table 1 describes the results on a cropped area of the *ROSIS Pavia University* and *Moffett Field* datasets. The three methods have comparable performances, with the proposed method being slightly better. Furthermore, the results strongly support the hypothesis that the target image has the same spatial structure as the panchromatic image used to train the model, otherwise keeping the posterior mixture weights would not lead to state-of-the-art results. Figures 2a to 2d illustrate the results from experiment 4 on the Moffett Field dataset, in terms of visual quality, yet the differences are not noticeable to the naked eye. Another way to visualize the difference in performance is to plot the (sorted) errors for each pixel and to notice

that the proposed method has a larger number below a given threshold, *i.e.*, in Figure 2e the dashed curve is below the solid curve, and furthermore, in Figures 2f and 2g the dashed curve is closer (smaller error) to the dotted one, representing the true pixel value across all bands, than the solid line.



**Fig. 2:** (a) Original HS bands in false color (20, 11, 4); (b) low-resolution; (c) dictionary-based [18]; (d) GMM-based (proposed); (e) Sorted pixel errors; (f)-(g) pixel error across bands.

## 7. CONCLUSION

This paper leveraged from the flexibility provided by the plug-and-play framework to address a data fusion problem known as hyperspectral image sharpening. For this specific problem, and motivated by recent results using class-adapted GMM priors on several image restoration/reconstruction problems [14, 15], we introduce what we called scene-adapted priors. As the name suggests, the model is learned from an image over the same scene we aim to reconstruct.

In addition, we propose some modifications for the GMM-based denoiser that allow us to show that the PnP-ADMM algorithm is guaranteed to converge to a global minimum of a cost function, defined by the choice of the denoiser itself. Experimental results show that the proposed method outperforms another state-of-the-art algorithm based on sparse representations on learned dictionaries [18], for most of the test settings. Moreover, the modifications also improve the results over the GMM-based denoiser used in [16], giving supporting evidence to the hypothesis that the panchromatic image constitutes a very useful training example, as the hyperspectral bands (and corresponding latent images) share a similar spatial structure.

## 8. REFERENCES

- [1] M. Afonso, J. Bioucas-Dias, M. Figueiredo, “An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems”, *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.
- [2] H. Bauschke, P. Combettes, “Convex Analysis and Monotone Operator Theory in Hilbert Spaces”, *Springer*, 2011.
- [3] R. Bhatia, “Matrix Analysis”, *Springer*, 1997.
- [4] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, “Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches”, *IEEE Jour. Sel. Topics in Appl. Earth Obser. and Remote Sens.*, vol. 5, pp. 354–379, 2012.
- [5] J. Bioucas-Dias and J. Nascimento “Hyperspectral subspace identification,” *IEEE Trans. Geosc. Remote Sens.*, vol. 46, pp. 2435–2445, 2008.
- [6] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2011.
- [7] S. Chan, X. Wang, O. Elgandy, “Plug-and-Play ADMM for Image Restoration: Fixed Point Convergence and Applications”, *IEEE Trans. Computational Imaging*, vol. PP, no. 99, pp. 1–1, 2016
- [8] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, “Image denoising by sparse 3D transform-domain collaborative filtering”, *IEEE Trans. Image Processing*, vol. 16, pp. 2080–2095, 2007.
- [9] L. Loncan, L. Almeida, J. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simões, J.-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei, N. Yokoya, “Hyperspectral pansharpening: a review”, *IEEE Geosc. Remote Sens. Mag.*, vol. 3, pp. 27–46, 2015.
- [10] M. Simões, J. Bioucas-Dias, L. Almeida, J. Chanussot, “A convex formulation for hyperspectral image superresolution via subspace-based regularization”, *IEEE Trans. Geosc. Remote Sens.*, vol. 55, pp. 3373–3388, 2015.
- [11] S. Sreehari, S. Venkatakrishnan, B. Wohlberg, G. Buzzard, L. Drummy, J. Simmons, C. Bouman “Plug-and-play priors for bright field electron tomography and sparse interpolation”, *IEEE Trans. Computational Imaging*, vol. 2, no. 4, pp. 408–423, 2016.
- [12] J. Sulam, Y. Romano, M. Elad “Gaussian mixture diffusion”, *IEEE Int. Conf. on the Science of Electrical Engineering*, 2016.
- [13] A. Teodoro, M. Almeida, M. Figueiredo, “Single-frame image denoising and inpainting using Gaussian mixtures”, *Int. Conf. on Pattern Recognition Applications and Methods*, pp. 283–288, 2015.
- [14] A. Teodoro, J. Bioucas-Dias, M. Figueiredo, “Image restoration and reconstruction using variable splitting and class-adapted image priors,” *IEEE Int. Conf. on Image Processing*, 2016.
- [15] A. Teodoro, J. Bioucas-Dias, M. Figueiredo, “Image restoration with locally selected class-adapted models,” *IEEE Int. Work. on Machine Learning for Signal Processing*, 2016.
- [16] A. Teodoro, J. Bioucas-Dias, M. Figueiredo. “Sharpening hyperspectral images using plug-and-play priors,” *Int. Conf. on Latent Variable Analysis and Signal Separation*, 2017.
- [17] S. Venkatakrishnan, C. Bouman, E. Chu, B. Wohlberg, “Plug-and-play priors for model based reconstruction”, *IEEE GlobalSIP*, pp. 945–948, 2013.
- [18] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, “Hyperspectral and multispectral image fusion based on a sparse representation”, *IEEE Trans. Geosc. Remote Sens.*, vol. 53 pp. 3658–3668, 2015.
- [19] G. Yu, G. Sapiro, and S. Mallat, “Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [20] D. Zoran, Y. Weiss, “From learning models of natural image patches to whole image restoration”, *IEEE Int. Conf. on Computer Vision and Pattern Recognition* pp. 479–486, 2011.